EPA Basic KM TEQ and ISM UCL Calculator

General Instructions: Password to protect/unprotect worksheets = "dioxin'

These instructions apply to the basic Kaplan-Meier (KM) Toxicity Equivalent (TEQ) calculator, which includes calculations that support a simple calculation of TEQs from congener concentrations and a determination of the **TEQ upper confidence limit (UCL) for a decision unit (DU) based on incremental sampling**. An Advanced Kaplan-Meier (KM) Toxicity Equivalent (TEQ) calculator is also available. It includes calculations that support a simple, quasi-sensitivity analysis that examines the effect of various ways of handling nondetected (ND) or rejected (R-flagged) analytical data results within a sample's congener profile.

BE AWARE

Individual statisticians vary in their acceptance of Helsel's adaptation of the Kaplan-Meier (KM) technique to estimate sample TEQs when nondetected congeners are present. (More details of this technique are covered in the "KM discussion" worksheet.) Other methods to avoid simple substitution for nondetects were suggested by peer reviewers of this calculator, and they may be incorporated into future updates of this calculator. The user is advised to seek input from a qualified statistician if important project or site decisions are dependent upon the choice of TEQ calculation method.

Helsel, D.R. 2009. "Summing Nondetects: Incorporating Low-Level Contaminants Risk Assessment." *Integrated Environmental Assessment and Management.* Volume 6, Number 3. Pages 361 through 366.

These instructions describe the iterative mechanics of using the worksheets in the workbook to compute TEQs from congener concentrations and determining the TEQ UCL for a DU.

This technique is a modification of the KM method for dealing with nondetect (ND) data in analytical data sets [(1)]. Slight modification allows it to be used to calculate the TEQ for a congener profile containing NDs. The user should refer to the write-up in the Uniform Federal Policy for Quality Assurance Project Plans (UFP-QAPP) Template User Guide discussing the handling of ND and rejected data (http://www.epa.gov/superfund/health/contaminants/dioxin/pdfs/Dioxin%20UFP%20QAPP%20UserGuide.pdf). Note that this TEQ calculation worksheet process also works for congener profiles that have detects for all congeners. An advanced version of the calculator is available for conducting sensitivity analysis for samples with rejected data or NDs that represent the highest toxic equivalent concentration (TEC).

(1) Helsel, D.R. (2012), *Statistics for Censored Environmental Data Using Minitab and R*, 2nd ed. Wiley and Sons. 324 pp.

Unprotect worksheets

The workbook uses an automated macro to perform the calculations and provide error messages if necessary, allowing the user to correct and repeat the process until the data are correctly entered.
The worksheets in this workbook are protected to avoid unintentional change to equations. To modify the workbook, use the password "dioxin" to unprotect worksheets. Unprotecting the sheet can be performed using the Home/Format/Protection/Unprotect Sheet option. The protection will be re-enabled automatically each time the macro is run, so it is not necessary for the user to manually reenable protection.

Data received from the laboratory should be entered into the "Data entry & TEQ results" sheet. Data should be grouped such that each different population (as hypothesized in the conceptual site model (CSM)) should be entered into a separate workbook. This is because the worksheet will provide additional calculations (TEQ UCLs for DUs with only 1 TEQ result) based on the assumption that all samples involved in the calculations represent the same contaminant population. In other words, all DUs entered into a single workbook have approximately the same TEQ values, variability in dioxin concentration across the DU, and congener profiles. Those samples that are determined to be from a different population should be removed from the workbook if the other calculations (UCLs, etc.) available in the worksheet are to be used.

Inserting additional sample rows

The worksheet is set up to accommodate up to 20 samples. If more than 20 samples are required, samples can be added by unprotecting the worksheet (password is "dioxin") and then adding the needed rows above the 20th sample. The three rows that are associated with an existing sample result should be copied onto three new rows. Rows can also be deleted if desired, but it is preferable to leave the data fields blank (including the sample identification (ID) in column B). Any blank rows should be inserted above the last sample row at bottom of the worksheet. If more triplicate rows are necessary, the rows at the top of the template can be copied over other sample rows, or at the end (be sure to insert 9 new rows before doing this). Again, the worksheet will need to be unprotected first. Note that when inserting, deleting or copying rows, it is possible to create program errors, so avoid this if possible.

Hiding columns

Columns should not be deleted from the worksheet. If congeners are not needed (not reported in the data set), the cells should be left blank. For ease of use, the unused columns can be hidden. The user will have to unprotect the worksheet to hide or unhide columns. Columns should not be deleted from the worksheet. Hide columns using the "Home" menu: Format/Visibility/Hide & Unhide/Hide Columns.

Excel format

Note that the calculator workbook is saved in Excel 97-2003 Workbook format (*.xls). The workbook should work properly in Excel 2007 and Excel 2010, and may be saved in Excel Macro-Enabled Workbook format (*.xlsm). In Excel 2007 and 2010 versions, the Excel Workbook format (*.xlsx) will not allow the macros in the calculator to operate properly, and should not be used to save the workbook unless all data processing is complete.

Instructions for Macro Use:

Step	Instruction
Enable macros	**Note: Prior to their use, the macros will first need to be enabled. In Excel 2007, this can be done by selecting 'Options' on the Security Warning bar that appears below the Excel menu bars when the workbook is opened, and selecting the 'Enable this content' button, then selecting the 'OK' button. For other versions of Excel, consult Excel HELP to determine how to enable macros.** Once the macros are enabled, follow the steps below.

1	Enter the sample numbers in column B of the "Data entry & TEQ results" worksheet. The sample numbers should be entered in the top row (Row A) of each three-row grouping. If a sample number in any Row A is left blank, the macro will stop operation after the previous sample and will not execute for any samples after this blank sample number.
2 Reorder congeners to match lab report	**Note:** This step is optional, but may help increase the speed and accuracy of manual data input. Check the order of the chemical names to ensure they are listed in the same order as the data reports planned to be used for data input. If they are not in the same order, change the numbers in row 5 so that they correspond to the order on project data reports. Then, click the button labeled "Sort Chemicals", which will run a macro to sort the analytes into the order specified in row 5. Note that the "Congener Abbreviations" worksheet contains a table listing the International Union of Pure and Applied Chemistry (IUPAC) names, Chemical Abstracts Service (CAS) numbers, and common abbreviations. This worksheet may be useful in matching the analyte names on the data reports to those in the data entry worksheet. After the sort is complete, check the worksheet again to ensure the chemical names are listed in the correct order. The step can be repeated as many times as necessary.
3 Entering qualifiers Pasting Data EMPC qualifier	Enter the congener data into Row A for each sample, along with any qualifiers (for example, J, E, A, ND, U) assigned to each result. The numeric portion of the result should be entered first, followed by the qualifier. It is not necessary to enter a space between the number and qualifier, but entering a space is also acceptable if the user prefers that approach. Row B will be automatically populated by the macro, so it is not necessary for the user to enter a result in this row. If the user wishes to copy and paste data into the spreadsheet, the Paste Values option should be used. To paste values, select "Paste" on the Excel ribbon, then "Paste Special", then "Paste As Values". Note that if estimated maximum possible concentration (EMPC) values are present, these values should be entered as nondetects (U or ND) with the EMPC value as the detection limit. This will ensure that these values are subjected to the full sensitivity analysis as nondetects with a maximum value of the EMPC. Also see the EMPC discussion in the "ND&R discussion" worksheet.
4	Row C (red line) is pre-programmed to immediately calculate the TEC for each congener from the numerical entries in Row B.
5 "Donor" samples	Click on either of the two boxes labeled "Calculate KM TEQ". This will run macros that will copy the entered data to the KM congener intermediate worksheet, and display the returned results. If a message indicating that data entry is incomplete is displayed, check that data have been entered in all cells in the data entry range. **If an "Error, see instructions." message appears, the largest TEC will be associated with an ND. This is a complication because the highest TEC value in the KM worksheet must be a detected result for the method to be valid. There are 2 options for dealing with this situation:** • Option 1) Process the data as usual, except that the first entry in the KM data column is marked as a detect even though it is undetected (U) or nondetected (ND); or • Option 2) Examine the rest of the data set and look for samples with a congener profile and concentrations very similar to the problematic one. These similar samples are referred to as "donor" samples. See if the problem congener is detected in the "donor" samples. o If the problem congener is detected in any of the "donor" samples, evaluate whether a substitution of the detect value can defensibly be made for the U/ND. If there is more than one "donor" sample that could be used to provide a substitute for the U/ND, use the most conservative (i.e., highest) one. Note that the detected value must be less than the ND value for defensible substitution. o In cases where there is more than one possible donor value, the user should use the Advanced KM Calculator to perform a sensitivity analysis for this sample. Once the Advanced KM Calculator has identified a defensible approach to calculating the TEQ for the sample, the result can be entered into the Basic Calculator. o Enter the designated donor value into Row A of the problem sample, remove the U or ND from Row A of the problem sample, and process through the KM worksheet as usual, treating the substitution as a detect. Make a note in column A for the problem sample that "the highest TEC was a ND value". Then record whether or not substitution was used to calculate the KM TEQ. If substitution was used, note from which sample ID the substituted value was obtained.
6	The KM intermediate mean value in the "KM congener intermediate" worksheet is automatically multiplied by the number of congeners that are part of the TEQ calculation. This calculation produces the TEQ value for that sample, which will be displayed in column AG ("Sample Total TEQ").
7	Examine the total TEQ values (column AG) and the congener profiles (columns C through AE). If they are all similar, they can be treated as a single population for the purpose of subsequent, potentially useful, calculations. Any samples that appear to belong to a different population should be moved to a different workbook. Alternatively, the outlying sample result can be deleted from the "Sample total TEQ" column since all subsequent calculations performed by the "Data entry & TEQ results" worksheet are performed on the values in that column (see the example workbook, a separate copy of the calculator).
8	All the TEQ results appearing in the "Sample total TEQ" column are reproduced in an uninterrupted column beginning in cell AH78 of the "Data entry & TEQ results" worksheet . This column can be copied and pasted using "paste special: values" into other worksheets, ProUCL, or other software.
9	It is recommended that the TEQ data be entered into ProUCL and data exploration be performed. This involves having ProUCL plot the data (using box plots, quantile-quantile (Q-Q) plots and histograms) to examine the data distribution for the group of DUs being treated as a single population in the CSM. These plots can be used to detect outliers, i.e., DUs that do not belong to the population.

	It is recommended that the graphs produced from ProUCL (or other software) be copied into the workbook as documentation of data assessment. See instructions below for "Instructions for using ProUCL with the calculator".

Instructions for using ProUCL with the calculator.	
1	The ProUCL User Guide should be consulted for how to use these plots/graphs to assess DU populations and data distributions (refer to p. 11 in the 4.00.05 version). Based on data exploration along with congener profiles, refine the TEQ data set and the CSM so that only data from the same population will be grouped together for further statistical analysis.
2	Once the data set has been refined so that all data are from the same population, statistical analysis can be performed using the rest of the automated UCL calculations in the "Data entry & TEQ results" worksheet.

3	Calculating a UCL for a DU requires a standard deviation (SD) that measures or estimates the variability across the DU. If three or more replicates are collected from a DU, then the UCL for that DU can be calculated directly. If only one composite sample is collected from a DU, there are indirect options for obtaining a SD to use. These options are described in the "Data entry & TEQ results" worksheet beginning with cell C75 ("Calculating TEQ UCLs for those DUs with only a single ICS result, and selecting the standard deviation value").
4	When the SD is selected, it can be entered into cell AN16. The n associated with the SD is entered into cell AP16. Enter an explanation of where the values came from in cell AT16.
5	The worksheet will calculate all two potential UCLs. The user must select the appropriate UCL from the dropdown box in cell (AS18) as documentation of the choice, and to have the correct UCLs appear in column AT.

Instructions for calculating a UCL TEQ for Replicate Samples	
1	Equations for calculating potential UCLs on three or more samples are also provided in the "Calc UCL from all DUs" worksheet. This worksheet allows the user to enter TEQ results that have not been calculated using the congener calculation process in the KM calculator.
2	There are two different UCL equations (Student's t and Chebyshev), and recommendations on which equation should be used are included on the "Calc UCL from all DUs" worksheet.

Note 1 Qualifiers	Note regarding sample qualifiers for KM TEQ results: All KM calculations include a determination of the TEC contribution to the TEQ from congener results that are qualified as nondetect U or ND) or estimated (J or E). Rejected (R-qualified) values are best addressed in the Advanced KM TEQ Calculator. If the contribution of these "qualified" TECs to the TEQ is greater than 50 percent, the KM TEQ result should be qualified. The qualifier is determined by the macro, and is shown in column AH.
Note 2 Adjust TEFs	Note regarding toxic equivalence factors (TEF): The TEFs used in the calculator are from the World Health Organization (WHO) 2005 report. If necessary, the user can change the TEF values to earlier values, or updated values if they are available. The TEFs can also be adjusted for additional sensitivity analysis if desired. To update the TEFs, the user should unprotect the workbook, change the TEFs of concern and then rerun the macro.
Note 3	Note regarding number of detected congeners: There must be at least 3 detected congeners for the methodology in the KM TEQ calculator to be meaningful. If fewer thanthree detected congeners are present in the results for a sample entered into the calculator, an error message will be displayed to the user. No KM TEQ calculations will be conducted for that sample. "Not calculated" will be displayed in column AN, and a note will be displayed in column AZ stating that fewer than three detected results were present. For discussion, refer to the worksheet "ND&R discussion" under "Treatment of Nondetected Congeners."
Note 4 D/F vs PCB contributions to TEQ	Note regarding dioxin/furan contributions to sample TEQ: In columns AJ and AK, the percentage of TEQ contributed from dioxins and furans (column AJ) and dioxin-like PCBs (column AK) is shown.

For questions about this Calculator, contact Deana Crumbling at USEPA, crumbling.deana@epa.gov or (703) 603-0643.

Only the information below that is related to nondetect and EMPC-quali

This material is reproduced from the discussion (27Sep10 version) presented in Appendix 4 of the dioxi

Appendix 4: Calculation of Total Dioxin TEQs with Nondetect and Rejected Congeners

Helsel's Kaplan-Meier Approach

Calculation of sums or totals for multi-constituent chemicals [e.g., total dioxin TEQs, total PCBs, total polycyclic aromatic hydrocarbons (PAHs), etc.] has typically involved simple substitution of zero, one-half the detection limit (DL), or the DL for left-censored (nondetect or less-than values) congeners. Because this practice introduces bias to estimates used in statistical calculations, however, many sources now strongly recommend against the use of arbitrary surrogate values for nondetects (Helsel 1990, 2005a, 2005b, 2009; EPA 2006, 2009a, 2009b).

Helsel (2009) describes an approach for calculating totals using the KM product limit estimator, which is based on the following relationship between the "mean" of the toxic equivalence concentrations (TECs) and total TEQ for samples containing multiple congeners:

total concentration = "mean" TEC x n (where n is the number of congeners)

Note that this "mean" TEC is an intermediate value in the calculation that has no relationship to a mean TEQ for replicate DU samples. The KM estimator is a nonparametric maximum likelihood estimator that has been widely used in survival and failure analysis for more than 50 years (Kaplan and Meier 1958, Klein and Moeschberger 2003, Meeker and Escobar 1998). The KM estimator has only recently come into use in environmental assessment studies (Helsel 2005a), and is currently a default method used in EPA's ProUCL software for calculating the 95% UCL of the mean for data with one or more censored results (EPA 2009a, 2009b).

Treatment of Nondetected Congeners

For the purposes of this dioxin reassessment UFP-QAPP template, the intermediate KM "mean" is recommended for use in calculating total dioxin TEQs, using the general equation presented above, in all cases where a) some fraction of the congeners are nondetect, and b) there are at least three detected congeners. Additional guidelines for calculating the KM intermediate "mean" are provided below. If all congeners are detected, then the intermediate "mean" calculated by the equation is the arithmetic average of all the congeners' TECs.

If only one or two congeners are detected, then there is no statistically satisfactory method for calculating the dioxin TEQ that adequately accounts for the uncertainty introduced by nondetect congener results. In this case, the intermediate "mean" should be calculated as the arithmetical average, where simple substitution is used for nondetects. A quasi-sensitivity analysis approach is recommended, wherein substitution of both zero and the DL are used to calculate lower- and upper-bound estimates for the total TEQ. Compare the

recommended. Cases where only one or two congeners are detected are discussed above. Lastly, Helsel (2009) recommends that for left-censored environmental data, Efron's bias correction should always be used. This simply requires that the minimum result always be treated as a detected result. The manner in which Efron's bias correction is incorporated in calculations of the KM mean depends on the specific software or approach used. For example, for programs that require a "flag" to distinguish between detected and nondetect data, one only needs to use the appropriate flag for detected data to qualify the minimum result(s).

Three options are described below for calculation of the KM mean:

(1) Helsel's KM Excel spreadsheet model (available from www.practicalstats.com). This spreadsheet has been built into a workbook designed specifically for calculating the TEQ from raw data congener concentration data. Raw data are entered into one spreadsheet, which automatically calculates the toxic equivalent concentration (TEC) for each congener. The TECs are copied and pasted into a second spreadsheet in the workbook that performs the KM calculation. This produces an intermediate value (the KM "mean") which is transferred back to the first spreadsheet. The intermediate result is then automatically multiplied by the number of congeners to produce the total TEQ for the sample. Detailed instructions for using the spreadsheets are included in the Excel workbook's spreadsheets.

(2) Alternatively, EPA's ProUCL software may be used. Before estimates of the KM intermediate "mean" TEC can be calculated, the congener concentration results (in ppt) must be converted to congener TECs by multiplying each congener by its TEF. This must be done independently before the TECs are put into ProUCL for the KM calculation. (ProUCL cannot do the TEC calculation.) The TECs are then entered into ProUCL and the KM intermediate "mean" is automatically calculated for data sets with one or more nondetect results. EPA (2009a, 2009b) should be consulted for instructions for entering data into ProUCL, since a coding procedure must be used in ProUCL to "tell it" which congener TECs were from ND values. Note that in order to use Efron's bias correction, the minimum result should be coded as a detected result. If intermediate "means" are required for multiple samples, then each sample needs to be identified using a "grouping" variable (see EPA 2009a). For each sample, the KM intermediate "mean" will need to be extracted from the ProUCL report, and manually multiplied by the number of congeners to produce the total TEQ result for that sample.

(3) Commercial or other statistical software. The KM model is included in many mainstream statistical software packages, as well as public domain (including the R language) programs. Helsel (2005a) discusses an approach for "flipping" data for use in commercial packages, which emphasize treatment of right-censored data. Experienced users may elect to use alternative approaches for calculation of the KM intermediate "mean," but must use methods employing Efron's bias correction, and must demonstrate that results are comparable to the intermediate "means" calculated using Options (1) or (2) above.

elect to perform a quasi-sensitivity analysis by calculating TEQ without the EMPC values. As for rejected data, significant effects from EMPC values may require corrective action to improve data quality (such as sample reanalysis).

Therefore, for congeners that are influential (high-toxicity, TEF close to 1, or high concentration) in calculations of the intermediate "mean" and total TEQ, rejected and qualified data may require further evaluation by project teams. The uncertainty of calculating total TEQs, as can be demonstrated through sensitivity analyses, should be addressed in the uncertainty section of assessment documents, and taken into account in decision making.

fied congeners is relevant to this Basic TEQ calculator.

in reassesment UFP-QAPP User Guide.

TEQs from both approaches to assess whether they have the same decision outcome. Substitution of one-half the DL can be used to calculate a "middle-of-the-road" value, although it should be acknowledged that the uncertainty of this estimate may be unacceptable for decision making.

In cases where critical decisions hinge on total TEQ estimates with mostly nondetect results, project teams are advised to consider

- consulting personnel with expertise in statistics,
- reanalyzing existing samples (if archived samples are available and meet holding times),
- comparing with results from nearby similar DUs and the CSM, or
- collecting additional samples.

The stepwise KM approach for calculating the total dioxin TEQ for individual samples is described below:

Step 1. Calculate the TEC for each congener by multiplying the results for individual congeners by their congener-specific TEF (van den Berg and others 2006). For nondetect congeners, the reporting limit or DL should be multiplied by the TEF.

Step 2. Calculate the intermediate "mean" TEC for each sample using a KM calculator spreadsheet. If all the congeners are detected, then calculate the intermediate value as the arithmetic mean. If nondetects are present and at least three results are detected, calculate the KM intermediate using one of the options described below. If only one or two congeners are detected, use simple substitution and a quasi-sensitivity analysis approach, as discussed above.

Step 3. Calculate the total dioxin TEQ using: Total TEQ = intermediate "mean" TEC x n, where n is the number of congeners in the calculation.

Helsel (2009) discusses several potential contraindications for calculation of the KM mean. The first concerns cases where only a single DL is used for all nondetect congeners. This is not expected to occur for calculation of total dioxin TEQs, since results for individual congers are first multiplied by congener-specific TEFs. The second contraindication is when the maximum reported result is a nondetect, high-toxicity (i.e., TEF close to 1) congener. This is problematic, as the KM method will effectively ignore maximum results that are censored. Helsel (2009) suggests that the DL be substituted in these cases, but that it should be acknowledged that this represents a worst-case scenario. Another option is to compare the congener concentration and congener profile of the sample with a high TEF nondetect to results from similar (per the CSM) DUs. If the congener profiles are similar, but the other DUs have a detection for the congener in question, substitution of a value (straight substitution, an average of several, or a maximum) from the other DUs may be made.

Helsel (2009) does not discuss the minimum number of detected results required to estimate the KM mean, but a practical minimum of three detected results is

Treatment of R-Qualified Congeners

One additional component for assessing the uncertainty of estimates of the intermediate KM "mean" and total TEQ, concerns treatment of rejected (R qualified) data. It is possible to reject individual congener analytes based on ion abundance, the signal-to-noise ratio, relative retention time, a low laboratory control sample result, gross blank contamination, or other analyte-specific criteria. For non-dioxin individual chemicals with multiple-sample sample sets (i.e., sufficient sample-sizes to support calculations), rejected data are always excluded from calculations in environmental assessments. However, for calculation of the "mean" (and total) for a set of congeners, there is concern that exclusion of rejected data may bias estimates low or create a need for replacement data (resampling or reanalysis). The magnitude (and importance) of this bias will of course depend on the values reported for R-qualified data, as well as the congener-specific TEFs.

Although rejected data should not be included in final calculations of TEQ for a given sampling or decision unit, rejected data values (concentrations or detection limits) can be included in KM "mean" and total TEQ calculations early in the data evaluation process. These TEQs can be compared to TEQs calculated with the rejected values removed. This quasi-sensitivity approach, similar to that recommended above for nondetect values, will assist project teams in assessing the magnitude of impacts from rejected data and the need for replacement data (Replacement data may require reanalysis of samples at the laboratory, with laboratory corrective actions or method refinements as needed, or the collection of additional samples from the site). Rejected data can be further evaluated through professional judgment, such as whether a rejected congener may be present at a concentration that could affect the TEQ based on historical site information or data from surrounding decision units. For example, project teams could use the KM calculator to further assess how high the concentration of a rejected congener would have to be to affect the TEQ, and then compare this estimate to concentrations for this congener that are present in other decision units, or in comparable historical data sets.

Treatment of EMPC values and qualified data

The CLP SOW for dioxin analysis specifies the reporting of detected congeners as "EMPC" values ("estimated maximum possible concentration") when a congener peak is present at an acceptable signal-to-noise ratio, but ion abundance criteria are not met for definitive identification of that congener. The CLP SOW excludes these values from the calculation of TEQ. EPA Method 8290A also specifies the reporting of EMPC values but makes no recommendations concerning their use in TEQ calculations. EMPC values are generally qualified as estimated concentrations ("J") or nondetect values ("U") during data validation in accordance with EPA Functional Guidelines. When qualified "J", EMPC values can be applied along with other J-qualified congener results in TEQ calculation and risk assessment (J-qualified data are generally applied like unqualified data under EPA risk assessment protocols). EMPC values qualified "U" can be treated as other nondetect values using the KM approach described above. Given that use of EMPC values may overestimate the TEQ and associated dioxin risk, project teams may again

References

Helsel, D.R. 1990. Less than obvious, statistical treatment of data below the detection limit. *Environmental Science and Technology*. Volume 24, Number 12. Pages

1767 through 1774.

Helsel, D.R. 2005a. *Nondetects and Data Analysis*. Statistics for Censored Environmental Data. John Wiley and Sons, Inc. Hoboken, NJ. 250 pages.

Helsel, D.R. 2005b. "More than Obvious: Better Methods for Interpreting Nondetect Data." *Environmental Science and Technology*. Volume 39, Number 20. Pages 419A through 423A.

Helsel, D.R. 2009. "Summing Nondetects: Incorporating Low-Level Contaminants in Risk Assessment." *Integrated Environmental Assessment and Management*. Volume 6, Number 3. Pages 361 through 366.

Kaplan, E.L., and P. Meier. 1958. "Nonparametric Estimator from Incomplete Observations." *Journal of the American Statistical Association*. Volume 53. Pages 457 through 481.

Klein, J.P., and M.L. Moeschberger. 2003. Survival Analysis – Techniques for Censored and Truncated Data. 2nd Edition. Statistics for Biology and Health, Springer Verlag. 560 pages.

Meeker, W.Q., and L.A. Escobar. 1998. Statistical Methods for Reliability Data. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc. New York, New York. 712 pages.

U.S. Environmental Protection Agency (EPA). 2006. "On the Computation of a 95% Upper Confidence Limit of the Unknown Population Mean Based Upon Data Sets with Below Detection Limit Observations." Singh, A., Maichle, R., and S.E. Lee. EPA/600/R-06/022. March.

EPA. 2009a. "ProUCL Version 4.00.04 Technical Guide (Draft)." Prepared by A. Singh and A.K. Singh. EPA/600/R-07/041. February.

EPA. 2009b. "ProUCL Version 4.00.04 User Guide (Draft)." Singh, A., R. Maichle, A.K. Singh, S.E. Lee, and N. Armbya. Office of Research and Development, National Exposure Research Laboratory. EPA/600/R-07/038. February.

Van den Berg, M. and others. (2006). "The 2005 World Health Organization Reevaluation of Human and Mammalian Toxic Equivalency Factors for Dioxins and Dioxin-Like Compounds." *Toxicological Sciences*. Volume 93, Number 2. Pages 223 through 241. On-Line Address: http://epa-prgs.ornl.gov/chemicals/help/documents/vandenberg2006.pdf

EPA ISM Basic KM TEQ and ISM UCL Calculator

Project Name:
Matrix: **Data entered by:**
Units: **Date entered:**

protect/unprotect sheet password = dioxin

	Chemical Sort Order	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29					
	WHO 2005 TEFs =	**1**	**1**	**0.1**	**0.1**	**0.1**	**0.01**	**0.0003**	**0.1**	**0.03**	**0.3**	**0.1**	**0.1**	**0.1**	**0.1**	**0.01**	**0.01**	**0.0003**	**0.0001**	**0.0003**	**0.00003**	**0.00003**	**0.00003**	**0.00003**	**0.1**	**0.00003**	**0.00003**	**0.00003**	**0.03**	**0.00003**	**KM Method**	**Qualifiers and Qualifier Fractions**			
Sample Notes	**Sample ID** **(must enter on Row A)**	**TCDD**	**PeCDD**	**1,6-HxCDD**	**1,4-HxCDD**	**1,9-HxCDD**	**1,4,6-HpCDD**	**OCDD**	**TCDF**	**1-PeCDF**	**4-PeCDF**	**1,4-HxCDF**	**1,6-HxCDF**	**1,9-HxCDF**	**4,6-HxCDF**	**1,4,6-HpCDF**	**1,4,9-HpCDF**	**OCDF**	**PCB 77**	**PCB 81**	**PCB 105**	**PCB 114**	**PCB 118**	**PCB 123**	**PCB 126**	**PCB 156**	**PCB 157**	**PCB 167**	**PCB 169**	**PCB 189**	**Sample total TEQ**	**TEQ qualifier**	**Fraction from nondetect and estimated results**	**Fraction from dioxins and furans**	**Fraction from dioxin-like PCB congeners**
Use this yellow block of 3 samples for triplicate ICSs in a single DU	value to use for KM: Row B																																		
	TEC for KM: Row C	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0.0000				
	value to use for KM: Row B																																		
	TEC for KM: Row C	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0.0000				
	value to use for KM: Row B																																		
	TEC for KM: Row C	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0.0000				
	value to use for KM: Row B																																		
	TEC for KM: Row C	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0.0000				
	value to use for KM: Row B																																		
	TEC for KM: Row C	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0.0000				
	value to use for KM: Row B																																		
	TEC for KM: Row C	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0.0000				
	value to use for KM: Row B																																		
	TEC for KM: Row C	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0.0000				
	value to use for KM: Row B																																		
	TEC for KM: Row C	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0.0000				
	value to use for KM: Row B																																		
	TEC for KM: Row C	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0.0000				
	value to use for KM: Row B																																		
	TEC for KM: Row C	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0.0000				
	value to use for KM: Row B																																		
	TEC for KM: Row C	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0.0000				
	value to use for KM: Row B																																		
	TEC for KM: Row C	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0.0000				
	value to use for KM: Row B																																		
	TEC for KM: Row C	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0.0000				
	value to use for KM: Row B																																		
	TEC for KM: Row C	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0.0000				
	value to use for KM: Row B																																		
	TEC for KM: Row C	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0.0000				
	value to use for KM: Row B																																		
	TEC for KM: Row C	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0.0000				
	value to use for KM: Row B																																		
	TEC for KM: Row C	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0.0000				
	value to use for KM: Row B																																		
	TEC for KM: Row C	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0.0000				
	value to use for KM: Row B																																		
	TEC for KM: Row C	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0.0000				
	value to use for KM: Row B																																		
	TEC for KM: Row C	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	0.0000				

	triplct SD	triplct RSD	1-sided UCL95(t)	Chebyshev UCL95
0.0	0.0	#DIV/0!	0.0	0

See worksheet "Calc UCL from all DUs" for explanation of which UCL to choose.

Enter SD used = 34.27228 **n used =** 20 From what were they selected? example only

DU TEQ UCL95(t)	DU TEQ UCL95(Cheb)	Select which UCL value to use for all samples in cell AS17
13.3	33.4	UCL95(t) 13.25123
DU TEQ UCL95(t)	DU TEQ UCL95(Cheb)	
13.3	33.4	UCL95(t) 13.25123
DU TEQ UCL95(t)	DU TEQ UCL95(Cheb)	
13.3	33.4	UCL95(t) 13.25123
DU TEQ UCL95(t)	DU TEQ UCL95(Cheb)	
13.3	33.4	UCL95(t) 13.25123
DU TEQ UCL95(t)	DU TEQ UCL95(Cheb)	
13.3	33.4	UCL95(t) 13.25123
DU TEQ UCL95(t)	DU TEQ UCL95(Cheb)	
13.3	33.4	UCL95(t) 13.25123
DU TEQ UCL95(t)	DU TEQ UCL95(Cheb)	
13.3	33.4	UCL95(t) 13.25123
DU TEQ UCL95(t)	DU TEQ UCL95(Cheb)	
13.3	33.4	UCL95(t) 13.25123
DU TEQ UCL95(t)	DU TEQ UCL95(Cheb)	
13.3	33.4	UCL95(t) 13.25123
DU TEQ UCL95(t)	DU TEQ UCL95(Cheb)	
13.3	33.4	UCL95(t) 13.25123
DU TEQ UCL95(t)	DU TEQ UCL95(Cheb)	
13.3	33.4	UCL95(t) 13.25123
DU TEQ UCL95(t)	DU TEQ UCL95(Cheb)	
13.3	33.4	UCL95(t) 13.25123
DU TEQ UCL95(t)	DU TEQ UCL95(Cheb)	
13.3	33.4	UCL95(t) 13.25123
DU TEQ UCL95(t)	DU TEQ UCL95(Cheb)	
13.3	33.4	UCL95(t) 13.25123
DU TEQ UCL95(t)	DU TEQ UCL95(Cheb)	
13.3	33.4	UCL95(t) 13.25123
DU TEQ UCL95(t)	DU TEQ UCL95(Cheb)	
13.3	33.4	UCL95(t) 13.25123
DU TEQ UCL95(t)	DU TEQ UCL95(Cheb)	
13.3	33.4	UCL95(t) 13.25123
DU TEQ UCL95(t)	DU TEQ UCL95(Cheb)	
13.3	33.4	UCL95(t) 13.25123

DATA LIST

This section copies the TEQ results above so they are easy to copy and paste into other spreadsheets in this workbook or ProUCL.
Note that any triplicate samples need to be corrected manually here - all 3 show up in the list.

Line	Sample ID	TEQ Result	Qualifier
line #1	1	2.1	
line #2			
line #3			
line #4			
line #5			
line #6			
line #7			
line #8			
line #9			
line #10			
line #11			
line #12			
line #13			
line #14			
line #15			
line #16			
line #17			
line #18			
line #19			
line #20			

Calculating TEQ UCLs for those DUs with only a single ICS result, and selecting the standard deviation value
If TEQ values & congener profile from pilot study DU similar to each DU results from the main investigation,
can use DU-P SD to calc DU-specific 95UCLs for TEQ.
OR
If TEQ values & congener profile from the QC (triplicate) DU(s) (in the main investigation) is similar to the other DU results,
can use the QC DU SD to calc DU-specific 95UCLs for TEQ. If >1 QC DU SD available, use average or highest.
OR
If there are >2 DUs in a very similar area, can use the between-DU TEQs to calculate a SD for the area,
then use that SD to calc DU-specific 95UCLs for each TEQ.
OR
Use average of DU-P TEQ SD, DU-triplicate SD from main investigation, and area-wide SD to calc DU-specific 95UCLs for TEQ.

Make sure all DU results used to determine SD for other DUs belong to the same population, i.e., have similar TEQ values and congener profiles to those of the single ICS DUs.

EPA Basic KM TEQ and ISM UCL Calculator

congener TEC	# Detects	# Nondetects	Quantiles
1	1	0	0.667
1	1	0	0.333
0.1	1	0	0.000
			0.000
			0.000

EPA Basic KM TEQ and ISM UCL Calculator

This spreadsheet serves to calculate a TEQ UCL for a DU when that DU has at least three replicate TEQ results
It includes an equation for calculating a 95% UCL using the Chebyshev non-parametric method.
The Chebyshev equation is from the ProUCL Technical Guide 4.00.04 (Feb 2009), p. 54, Eqn 2-46.
Unacceptably high standard deviations (and thus UCLs) are an indication that some aspect of data variability is not being controlled. The largest source(s) of variability should be found and corre
The 3 most common causes of problems are mixed populations, lax sample homogenization/subsampling, and too few increments.
Which UCL (t- or Chebyshev) should be used? See below for the answer from the ITRC ISM team statistician

Sample ID: triplicate example

Enter the replicate TEQ results here	21.69
	23.36
	23.36

(all the parameters below will automatically calculate)

n =	3
mean =	22.80
Std Dev =	0.96
RSD =	0.042

The 1-sided 95% UCL using the t-distribution:
UCL = 24.43

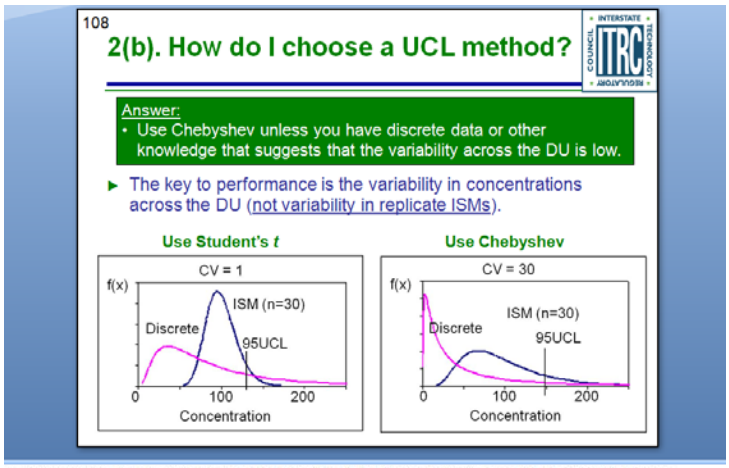
protect/unprotect sheet password = dioxin

The nonparametric Chebyshev 95% UCL:
UCL = 25.23

To calculate an RSD & t- & Chebyshev UCL from a large group of ICS results

Enter the TEQ results here:	(example)
1	22.80
2	2.10
3	50.783
4	14.24617
5	10.77325
6	15.72938
7	10.3186
8	5.231167
9	2.195317
10	15.08784
11	4.560439
12	3.97536
13	7.51685
14	10.2427
15	11.257
16	
17	
18	
19	
20	
21	
22	
23	

n =	15
mean =	12.45
Std Dev =	12.05
RSD =	0.968
1-sided 95% t-distribution UCL =	17.9
95% Chebyshev UCL =	26.0



In these examples, the true mean is meant to be 100. The pink distribution represents discrete data; the blue distribution represents ISM data. Note that the more skewed the data (pink line, right figure versus left), the more likely that estimates of the mean using ISM will be low (see blue line on right, were data is more skewed, compared with the left). This potential for underestimating the mean when data are positive skewed was mentioned in earlier slides. If variability in the underlying data is high, there is a good chance that the Student's t calculation may underperform, i.e., give a 95UCL value that is than the actual 95% confidence level concentration. That is why when the variability is known or suspected to be high, or is simply unknown, the Chebyshev method should be used.

- ► Both methods provide desired 95% coverage when variability is low
- ► Chebyshev has more consistent 95% coverage for medium and high variability
- ► Increasing *r* (>3) and n (>30) provides marginal improvement in coverage for Chebyshev, but no improvement for Student's-*t*

Which UCL to Use

- If have discrete data showing (or information justifying an assumption of) low variability across the DU, use Student's-*t* equation.
  - E.g., "uniform" deposition by air transport
- If have discrete data or information showing high variability, use the Chebyshev eqn.
  - E.g., Spill areas, disturbed areas
- If don't know how variable the concentrations are across the DU, use the Chebyshev.
- Reference: ITRC ISM1 doc, Section 4.2.2

Abbreviation 1	Abbreviation 2	IUPAC name	CAS #
TCDD	2,3,7,8-TCDD	2,3,7,8-Tetrachlorodibenzo-p-dioxin	1746-01-6
PeCDD	1,2,3,7,8-PeCDD	1,2,3,7,8-Pentachlorodibenzo-p-dioxin	40321-76-4
1,4-HxCDD	1,2,3,4,7,8-HxCDD	1,2,3,4,7,8-Hexachlorodibenzo-p-dioxin	39227-28-6
1,6-HxCDD	1,2,3,6,7,8-HxCDD	1,2,3,6,7,8-Hexachlorodibenzo-p-dioxin	57653-85-7
1,9-HxCDD	1,2,3,7,8,9-HxCDD	1,2,3,7,8,9-Hexachlorodibenzo-p-dioxin	19408-74-3
1,4,6-HpCDD	1,2,3,4,6,7,8-HpCDD	1,2,3,4,6,7,8-Heptachlorodibenzo-p-dioxin	35822-39-4
OCDD	1,2,3,4,5,6,7,8-OCDD	Octachlorodibenzo-p-dioxin	3268-87-9
TCDF	2,3,7,8-TCDF	2,3,7,8-Tetrachlorodibenzofuran	51207-31-9
1-PeCDF	1,2,3,7,8-PeCDF	1,2,3,7,8-Pentachlorodibenzofuran	57117-41-6
4-PeCDF	2,3,4,7,8-PeCDF	2,3,4,7,8-Pentachlorodibenzofuran	57117-31-4
1,4-HxCDF	1,2,3,4,7,8-HxCDF	1,2,3,4,7,8-Hexachlorodibenzofuran	70648-26-9
1,6-HxCDF	1,2,3,6,7,8-HxCDF	1,2,3,6,7,8-Hexachlorodibenzofuran	57117-44-9
1,9-HxCDF	1,2,3,7,8,9-HxCDF	1,2,3,7,8,9-Hexachlorodibenzofuran	72918-21-9
4,6-HxCDF	2,3,4,6,7,8-HxCDF	2,3,4,6,7,8-Hexachlorodibenzofuran	60851-34-5
1,4,6-HpCDF	1,2,3,4,6,7,8-HpCDF	1,2,3,4,6,7,8-Heptachlorodibenzofuran	67562-39-4
1,4,9-HpCDF	1,2,3,4,7,8,9-HpCDF	1,2,3,4,7,8,9-Heptachlorodibenzofuran	55673-89-7
OCDF	1,2,3,4,5,6,7,8-OCDF	Octachlorodibenzofuran	39001-02-0
PCB 77	3,3',4,4'-TCB	3,3',4,4'-Tetrachlorobiphenyl	32598-13-3
PCB 81	3,4,4',5-TCB	3,4,4',5-Tetrachlorobiphenyl	70362-50-4
PCB 105	2,3,3',4,4'-PeCB	2,3,3',4,4'-Pentachlorobiphenyl	32598-14-4
PCB 114	2,3,4,4',5-PeCB	2,3,4,4',5-Pentachlorobiphenyl	74472-37-0
PCB 118	2,3',4,4',5-PeCB	2,3',4,4',5-Pentachlorobiphenyl	31508-00-6
PCB 123	2,3',4,4',5'-PeCB	2,3',4,4',5'-Pentachlorobiphenyl	65510-44-3
PCB 126	3,3',4,4',5-PeCB	3,3',4,4',5-Pentachlorobiphenyl	57465-28-8
PCB 156	2,3,3',4,4',5-HxCB	2,3,3',4,4',5-Hexachlorobiphenyl	38380-08-4
PCB 157	2,3,3',4,4',5'-HxCB	2,3,3',4,4',5'-Hexachlorobiphenyl	69782-90-7
PCB 167	2,3',4,4',5,5'-HxCB	2,3',4,4',5,5'-Hexachlorobiphenyl	52663-72-6
PCB 169	3,3',4,4',5,5'-HxCB	3,3',4,4',5,5'-Hexachlorobiphenyl	32774-16-6
PCB 189	2,3,3',4,4',5,5'-HpCB	2,3,3',4,4',5,5'-Heptachlorobiphenyl	39635-31-9

Type
Dioxin
Dioxin
Dioxin
Dioxin
Dioxin
Dioxin
Dioxin
Furan
Furan
Furan
Furan
Furan
Furan
Furan
Furan
Furan
Furan
Dioxin-like PCB
Dioxin-like PCB
Dioxin-like PCB
Dioxin-like PCB
Dioxin-like PCB
Dioxin-like PCB
Dioxin-like PCB
Dioxin-like PCB
Dioxin-like PCB
Dioxin-like PCB
Dioxin-like PCB
Dioxin-like PCB

Acronym List	
A	Data qualifier used to indicate an estimated result.
CAS	Chemical Abstracts Service
CSM	Conceptual site model
CV	Coefficient of variation
DU	Decision unit
E	Data qualifier used to indicate an estimated result.
HpCDD	Heptachlorodibenzo(p)dioxin
HpCDF	Heptachlorodibenzofuran
HxCDD	Hexachlorodibenzo(p)dioxin
HxCDF	Hexachlorodibenzofuran
ICS	Incremental composite sample
ID	Identification
ISM	Incremental sampling methodology
ITRC	Interstate Technology and Regulatory Council
IUPAC	International Union of Pure and Applied Chemistry
J	Data qualifier used to indicate an estimated result.
KM	Kaplan-Meier
ND	Nondetect
OCDD	Octachlorodibenzo(p)dioxin
OCDF	Octachlorodibenzofuran
PCB	Polychlorinated biphenyl
PeCDD	Pentachlorodibenzo(p)dioxin
PeCDF	Pentachlorodibenzofuran
QC	Quality control
RSD	Relative standard deviation
SD	Standard deviation
TCDD	Tetrachlorodibenzo(p)dioxin
TCDF	Tetrachlorodibenzofuran
TEC	Toxic equivalent concentration
TEF	Toxic equivalence factor
TEQ	Toxic equivalent
U	Data qualifier used to indicate a nondetected result.
UCL	Upper confidence limit
UFP-QAPP	Uniform Federal Policy - Quality Assurance Project Plan
WHO	World Health Organization